Impact Measures for Gradual Argumentation Semantics

Caren Al Anaissy
CRIL, Univ. Artois CNRS & LIP6, Sorbonne University
Lens & Paris, France
alanaissy@cril.fr,caren.al-anaissy@lip6.fr

Jérôme Delobelle
Université Paris Cité, LIPADE, F-75006
Paris, France
jerome.delobelle@u-paris.fr

Srdjan Vesic
CRIL CNRS Univ. Artois
Lens, France
vesic@cril.fr

Bruno Yun
Universite Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Lumière Lyon 2, LIRIS, UMR5205, 69622
Villeurbanne, France
bruno.yun@univ-lyon1.fr

ABSTRACT

Argumentation is a formalism allowing to reason with contradictory information by modeling arguments and their interactions. There are now an increasing number of gradual semantics to compute argument strengths and impact measures that have emerged to facilitate the interpretation of their outcomes. An impact measure assesses, for each argument, the impact of other arguments on its score. In this paper, we refine an existing impact measure and introduce a new impact measure rooted in Shapley values. We introduce several principles to evaluate those two impact measures w.r.t. some well-known gradual semantics. Our analysis provides deeper insights into the measures' functionality and desirability.

KEYWORDS

Impact measures; Argumentation; Gradual semantics

ACM Reference Format:
Caren Al Anaissy, Jérôme Delobelle, Srdjan Vesic, and Bruno Yun. 2025. Impact Measures for Gradual Argumentation Semantics. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

Group decision-making and negotiation are critical components in modern multi-agent systems [25, 32]. Effective information exchange between agents is essential for achieving coordination and cooperation. Computational argumentation theory provides a structured framework for agents to articulate and justify their positions, enabling them to selectively share their goals and intentions throughout the negotiation process. Moreover, computational argumentation theory stands as an important domain of artificial intelligence (AI), especially in knowledge representation and reasoning. It is used by agents for inferring conclusions in decision making problems [8, 22, 30] and for resolving conflicts of opinion in persuasion and negotiation dialogues [6, 37]. It represents knowledge in *argumentation graphs* with arguments as nodes and a binary attack relation for conflicts between pieces of information. Different semantics can then be applied on those graphs to obtain rational conclusions. In this paper, we focus on *gradual semantics* which evaluate and score each argument w.r.t. "how much" it is attacked by other arguments.

Recently, explainable AI has garnered more attention for its ability to provide transparency and enhance the understandability of AI-based models and algorithms. Among the existing notions of explainability in the literature [29, 40], two notions seem pivotal to paving the way toward explaining the outcomes of argument evaluation: *Causality* [23] and *Feature Attribution* [28, 36]. Causality refers to the ability to explain the link between what is introduced as input and what results as output. Feature attribution means assessing the contribution of input features (e.g., the age or the blood pressure) to the output (e.g., the probability of getting sick) by assigning attribution scores to each feature, avoiding the need to explore the model's internal mechanisms.

Hence, we argue that – working toward making argumentation interpretable – it is essential that the evaluation of an argument can be explained by highlighting the attribution of each argument's interconnected network to its final evaluation. This is achieved by providing to the user an impact measure that returns the contribution of different arguments to the argument's final evaluation.

Several impact measures have been defined in the literature [16, 24, 26, 33, 41]. We believe that there is still the need to explore other impact measures that have not been yet considered for gradual semantics. For this reason, this paper addresses the need for enhanced impact measures, bringing modifications to existing functions while introducing novel ones. Our study navigates the intricate landscape by first revising the impact measure defined in [16]. Additionally, we define a complementary impact measure based on the Shapley Contribution Measure [5, 39]. These impact measures can be used to assess the impact of an argument (or set of arguments) on a given argument for any existing gradual semantics, clarifying and explaining their contributions on the score of another argument.

There are three contributions. First, the enhancement of an existing impact measure and the definition of a novel impact measure based on Shapley Contribution Measure. Second, the introduction of nine principles for evaluating each (impact, semantics) pair and a full analysis of two impact measures under some well-known

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

gradual semantics. Lastly, the implementation of an online prototype platform where users can input their argumentation graphs, compute the acceptability degrees for a gradual semantics, and obtain the output of our new impact measures.

In Section 2, we start by a comprehensive review of existing literature and methodologies and transition into the definition of impact measures (Section 3). We then analyze our findings through the defined principles by analyzing their satisfiability by different (impact, semantics) pairs (Section 4). We propose a case study of a debate in a city council, where the aim will be to evaluate, among other things, the impact of agents involved in the debate (Section 5). The paper concludes by summarizing the implications of our work and outlining future directions in this evolving landscape (Section 6).

2 PRELIMINARIES

An argumentation framework (AF) is $\mathbf{AS} = (\mathcal{A}, C)$, where $\mathcal{A}$ is a finite set of arguments and $C \subseteq \mathcal{A} \times \mathcal{A}$ is a set of attacks between arguments. The *set of all direct attackers* of $x \in \mathcal{A}$ will be denoted as $Att(x) = \{y \in \mathcal{A} \mid (y, x) \in C\}$. Given two AFs $\mathbf{AS} = (\mathcal{A}, C)$ and $\mathbf{AS}' = (\mathcal{A}', C')$, $\mathbf{AS} \oplus \mathbf{AS}'$ is the AF $(\mathcal{A} \cup \mathcal{A}', C \cup C')$. For any AF $\mathbf{AS} = (\mathcal{A}, C)$ and $X \subseteq \mathcal{A}$, $\mathbf{AS}|_X = (X, C \cap (X \times X))$. The *set of external attackers of* X is $Arg^-(X) = \{y \in \mathcal{A} \backslash X \mid \exists x \in X \; s.t. \; (y, x) \in C\}$. The *set of external attacks to* X is the set of attacks from an argument in $Arg^-(X)$ to an argument in X and is formally defined as follows: $Att^-(X) = \{(y, x) \in C \mid y \in Arg^-(X) \text{ and } x \in X\}$. For every $x, y \in \mathcal{A}$, a *path* from y to x is a sequence $\langle x_0, \ldots, x_n \rangle$ of arguments such that $x_0 = y$, $x_n = x$ and $\forall i$ s.t. $0 \leq i < n$, $(x_i, x_{i+1}) \in C$. The attack structure of $x \in \mathcal{A}$ is a set of arguments that contains x and all the (direct or indirect) attackers and defenders of x.

Definition 1 (Attack Structure). *Let* $\mathbf{AS} = (\mathcal{A}, C)$ *be an AF and* $x \in \mathcal{A}$. *The attack structure of* x *in* $\mathbf{AS}$ *is* $\mathtt{Str}_{\mathbf{AS}}(x) = \{x\} \cup \{y \in \mathcal{A} \mid$ *there is a path from* y *to* $x\}$.

Example 1. *Consider the AF* $\mathbf{AS} = (\mathcal{A}, C)$ *represented in Figure 3. We have* $Att(a_4) = \{a_3, a_5, a_8\}$, $Arg^-(\{a_8, a_{10}\}) = \{a_9\}$, $Att^-(\{a_8, a_{10}\}) = \{(a_9, a_8), (a_9, a_{10})\}$, $\mathtt{Str}_{\mathbf{AS}}(a_3) = \{a_1, a_2, a_3\}$ *and* $\mathtt{Str}_{\mathbf{AS}}(a_4) = \{a_1, a_2, a_3, a_4, a_5, a_6, a_8, a_9, a_{10}\}$.

The usual Dung's semantics [17] extract justifiable sets of arguments (called extensions) from the argumentation framework. Those semantics induces a two-levels acceptability of arguments (inside or outside of one or all extensions). Gradual semantics (and ranking-based semantics) have been proposed as a more fine-grained approach to argument acceptability [2, 12]. These semantics use a weighting to assign to each argument in the argumentation framework a score, called (acceptability) degree.

Definition 2 (Gradual semantics). *A gradual semantics is a function* σ *which associates to each* $\mathbf{AS} = (\mathcal{A}, C)$, *a weighting* $\sigma_{\mathbf{AS}} : \mathcal{A} \rightarrow [0, 1]$ *on* $\mathcal{A}$. $\sigma_{\mathbf{AS}}(a)$ *is called the degree of* a.

Let us now recall some well-known gradual semantics studied in the literature [7, 10, 31].

The h-categoriser semantics (Hbs) assigns a value to each argument by taking into account the sum of degrees of its attackers, which themselves take into account the degree of their attackers.

Definition 3 (h-categoriser). *The h-categoriser semantics is a gradual semantics* σ^{Hbs} *s.t. for any* $\mathbf{AS} = (\mathcal{A}, C)$ *and* $a \in \mathcal{A}$:

$$\sigma^{\mathsf{Hbs}}_{\mathbf{AS}}(a) = \frac{1}{1 + \Sigma_{b \in Att(a)} \sigma^{\mathsf{Hbs}}_{\mathbf{AS}}(b)}$$

The Card-based semantics (Car) favors the number of attackers over their quality. This semantics is based on a recursive function which assigns a score to each argument on the basis of the number of its direct attackers and their degrees.

Definition 4 (Card-based). *The Card-based semantics is a gradual semantics* σ^{Car} *s.t. for any* $\mathbf{AS} = (\mathcal{A}, C)$ *and* $a \in \mathcal{A}$:

$$\sigma^{\mathsf{Car}}_{\mathbf{AS}}(a) = \frac{1}{1 + |Att(a)| + \frac{\Sigma_{b \in Att(a)} \sigma^{\mathsf{Car}}_{\mathbf{AS}}(b)}{|Att(a)|}}$$

The Max-based semantics (Max) favors the quality of attackers over their number. The degree of an argument is based on the degree of its strongest direct attacker.

Definition 5 (Max-based). *The Max-based semantics is a gradual semantics* σ^{Max} *s.t. any* $\mathbf{AS} = (\mathcal{A}, C)$ *and* $a \in \mathcal{A}$:

$$\sigma^{\mathsf{Max}}_{\mathbf{AS}}(a) = \frac{1}{1 + \max_{b \in Att(a)} \sigma^{\mathsf{Max}}_{\mathbf{AS}}(b)}$$

The last gradual semantics we study is the counting semantics (CS) [34]. It assigns a value to each argument by counting the number of their respective attackers and defenders. An AF is considered a dialogue game between the proponents of a given argument x (i.e., the defenders of x) and the opponents of x (i.e., the attackers of x). Thus, the degree of an argument is greater if it has many arguments from proponents and few arguments from opponents. Formally, they convert a given AF into its adjacency matrix $M_{n \times n}$ (where n is the number of arguments). The matrix product of k copies of M, denoted by M^k, represents, for all the arguments in AF, the number of defenders (if k is even) or attackers (if k is odd) situated at the beginning of a path of length k. Finally, a normalization factor N (e.g., the matrix infinite norm) is applied to M to guarantee convergence, and a damping factor α is used to have a more refined treatment of different lengths of attackers and defenders (i.e., shorter attacker/defender lines are preferred).

Definition 6 (Counting model). *Let* $\mathbf{AS} = (\mathcal{A}, C)$ *be an argumentation framework with* $\mathcal{A} = \{x_1, \ldots, x_n\}$, $\alpha \in (0, 1)$ *be a damping factor and* $k \in \mathbb{N}$. *The n-dimensional column vector* v *over* $\mathcal{A}$ *at step* k *is defined by* $v^k_\alpha = \sum_{i=0}^{k} (-1)^i \alpha^i \tilde{M}^i \mathcal{I}$, *where* $\tilde{M} = M/N$ *is the normalized matrix with* N *as normalization factor and* $\mathcal{I}$ *as the n-dimensional column vector containing only 1s.*
The counting model of $\mathbf{AS}$ *is* $v_\alpha = \lim_{k \to +\infty} v^k_\alpha$. *The degree of* $x_i \in \mathcal{A}$ *is the* i^{th} *component of* v_α, *denoted by* $\sigma^{\mathsf{CS}}_{\mathbf{AS}}(x_i)$.

A more detailed definition and examples can be found in [15, 34].

Together with the introduction of these semantics, a number of properties have been defined for gradual semantics to evaluate their behaviors (see [7, 11] for an overview). Two of the most well-known are called Independence and Directionality. The former ensures that the calculation of the acceptability degrees in two disconnected AFs should be independent while the latter ensures that the acceptability degree of an argument should only rely on the arguments with a directed path to it.

Property 1 (Independence [7]). *A semantics σ satisfies Independence iff for any two AFs $\mathbf{AS} = (\mathcal{A}, C)$, $\mathbf{AS}' = (\mathcal{A}', C')$, where $\mathcal{A} \cap \mathcal{A}' = \emptyset$, for every $y \in \mathcal{A}$, $\sigma_{\mathbf{AS}}(y) = \sigma_{\mathbf{AS} \oplus \mathbf{AS}'}(y)$.*

Property 2 (Directionality [7]). *A semantics σ satisfies Directionality iff for any AF $\mathbf{AS} = (\mathcal{A}, C)$, $\mathbf{AS}' = (\mathcal{A}, C \cup \{(b, x)\})$, for every $y \in \mathcal{A}$ such that there is no path from x to y, $\sigma_{\mathbf{AS}}(y) = \sigma_{\mathbf{AS}'}(y)$.*

To our knowledge, the counting semantics is the only one not to satisfy these two properties. However, we have chosen to study it as it differs from the other gradual semantics studied (Hbs, Car and Max), which all satisfy Independence and Directionality.

Proposition 1. *The counting semantics does not satisfy the Independence and Directionality properties.*

3 IMPACT FOR GRADUAL SEMANTICS

An impact measure is a function that informs on how a set of arguments "impacts" the score of a specific argument. In this paper, it returns a value within the interval $[-1, 1]$, reflecting the impact of the set and its overall polarity (negative, positive or neutral).

Definition 7 (Impact measure). *Let $\mathbf{AS} = (\mathcal{A}, C)$ be an AF and σ be a gradual semantics. An impact measure Imp takes as input $\mathbf{AS}$ and σ and returns a function $\mathtt{Imp}^{\sigma}_{\mathbf{AS}} : 2^{\mathcal{A}} \times \mathcal{A} \rightarrow [-1, 1]$. For any $X \subseteq \mathcal{A}$, $y \in \mathcal{A}$, $\mathtt{Imp}^{\sigma}_{\mathbf{AS}}(X, y)$ is the impact of X on y (in $\mathbf{AS}$ w.r.t. semantics σ).*

In Subsection 3.1, we explain the drawbacks of the impact measure by Delobelle and Villata [16] and introduce a revised version. We also motivate and showcase a new Shapley-based impact measure in Subsection 3.2. Note that these two impact measures have been implemented and deployed in an online platform prototype, accessible via https://github.com/brunoyun/django-app.

3.1 Revised Version of Impact Measure from Delobelle and Villata

In [16], the impact of an argument (or a set of arguments) on a target argument can be measured by computing the difference of acceptability degree of the target argument when this element exists and when it is deleted. To capture this notion of deletion, two deletion operators need to be defined. The *argument deletion operator* $\ominus^{\mathcal{A}}$ aims to delete a set of arguments from a given argumentation framework. These changes have also a direct impact on the set of attacks because the attacks directly related to the deleted arguments (attacking as well as attacked) are automatically deleted too.[1] The *attack deletion operator* $\ominus^{C}$ focuses only on the removal of a set of attacks from the initial argumentation framework, thus keeping the same set of arguments.

Definition 8 (Deletion operator). *Let $\mathbf{AS} = (\mathcal{A}, C)$ be an AF, $X \subseteq \mathcal{A}$ be a set of arguments, $R \subseteq C$ be a set of attacks and $y \in \mathcal{A}$. The argument deletion operator $\ominus^{\mathcal{A}}$ is defined as $\mathbf{AS} \ominus^{\mathcal{A}}_{y} X = (\mathcal{A}', C')$, where*

- *$\mathcal{A}' = \mathcal{A} \backslash (X \backslash \{y\})$;*
- *$C' = \{(x, z) \in C \mid x \in \mathcal{A} \backslash X, z \in \mathcal{A} \backslash X\}$.*

The attack deletion operator $\ominus^{C}$ is defined as $\mathbf{AS} \ominus^{C} R = (\mathcal{A}, C'')$, where $C'' = C \backslash R$.

In [16], to compute the impact of any set of arguments X on an argument y, it is proposed to consider the degree of acceptability of y when the direct attackers of X are removed (i.e., when the arguments in X are the strongest) from which the degree of acceptability of y is deducted when all the arguments of X are removed.

Definition 9 ([16]). *Let $\mathbf{AS} = (\mathcal{A}, C)$ be an AF, $y \in \mathcal{A}$ and $X \subseteq \mathcal{A}$. Let σ be a gradual semantics. ImpDV is defined as follows:*

$$\mathtt{ImpDV}^{\sigma}_{\mathbf{AS}}(X, y) = \sigma_{\mathbf{AS} \ominus^{\mathcal{A}}_{y} Arg^{-}(X)}(y) - \sigma_{\mathbf{AS} \ominus^{\mathcal{A}}_{y} X}(y)$$

Although this definition works well in general, there are particular cases (e.g., when self-attacks are allowed or when the direct attackers of X impact y via other paths) where the result does not correspond to what is expected. E.g., if $\mathbf{AS}_s = (\{a\}, \{(a, a)\})$, applying Definition 9 with the h-categoriser semantics (any other semantics gives the same result), we obtain $\mathtt{ImpDV}^{\mathsf{Hbs}}_{\mathbf{AS}_s}(\{a\}, a) = 0$ as it is not possible to delete the argument whose impact we want to evaluate. This result can be considered counter-intuitive because the degree of a is not maximal and the only argument who can have an impact on its degree is a itself.

We propose a revised version of ImpDV which takes into account this case and other problems while retaining the idea of the proposed approach. Thus, instead of removing all the direct attackers from X (i.e., using $\ominus^{\mathcal{A}}$), we remove the direct attacks on each argument of X (i.e., using $\ominus^{C}$). Note that in section 4.1, we introduce the Impact Existence principle (Principle 9) that states that if the acceptability degree of an argument a is not null, then there must exist at least a set of arguments whose impact on a is not null. Although the impact measure from Delobelle and Villata (Definition 9) violates this principle, as shown above with the case of having a single argument a attacking itself, the revised version that we propose satisfies this principle, leading to more intuitive results.

Definition 10 (Revised version of ImpDV). *Let $\mathbf{AS} = (\mathcal{A}, C)$ be an AF, $y \in \mathcal{A}$ and $X \subseteq \mathcal{A}$. Let σ be a gradual semantics. We have $\mathtt{ImpDV}^{\sigma}_{\mathbf{AS}}(X, y) = \sigma_{\mathbf{AS} \ominus^{C} Att^{-}(X)}(y) - \sigma_{\mathbf{AS} \ominus^{\mathcal{A}}_{y} X}(y)$.*

The problem mentioned in the self-attacking argument example is solved with the revised version because the self-attack is removed from the right-hand side of the formula. Thus, we have $\sigma^{\mathsf{Hbs}}_{\mathbf{AS}_s \ominus^{C} \{\emptyset\}}(a) - \sigma^{\mathsf{Hbs}}_{\mathbf{AS}_s \ominus^{\mathcal{A}}_{a} \{a\}}(a) \simeq 0.618 - 1 = -0.382$. In the following sections, ImpDV will refer to that defined in Definition 10.

Theorem 2. *Let σ be a gradual semantics and $\mathbf{AS} = (\mathcal{A}, C)$ be an AF. $\mathtt{ImpDV}^{\sigma}_{\mathbf{AS}}$ is an impact measure.*

Example 2. *Let $\mathbf{AS}$ be the AF depicted in Figure 3. We display the values of the revised impact of a set X on an argument y in Table 1. Using the same examples of Table 1, the original impact measure of [16] would yield the same values with the exception of the impact of $\{a_1\}$ on a_4 with a value of 0. To understand this difference, we need to compute the score of a_4 when a_2 is removed (because $Arg^{-}(\{a_1\}) = \{a_2\}$). Since a_1 and a_2 are symmetrical in terms of attacks received and given (i.e. they attack each other and both attack*

[1] Note that for $\ominus^{\mathcal{A}}$, it is necessary to specify the argument y for which the degree will be measured, in order to avoid removing it from the AF if it belongs to the set of arguments whose impact on y is to be measured.

X	y	$Att^-(X)$	$\sigma^{\text{Hbs}}_{\text{AS}_1}(y)$	$\sigma^{\text{Hbs}}_{\text{AS}_2}(y)$	$\texttt{ImpDV}^{\text{Hbs}}_{\text{AS}}(X,y)$
$\{a_1\}$	a_4	$\{(a_2,a_1)\}$	0.397	0.382	0.015
$\{a_5\}$	a_4	$\{(a_6,a_5)\}$	0.326	0.484	-0.158
$\{a_8,a_{10}\}$	a_4	$\{(a_9,a_{10}),(a_9,a_8)\}$	0.339	0.514	-0.175
$\{a_9\}$	a_4	$\{(a_{10},a_9)\}$	0.409	0.339	0.07
$\{a_4\}$	a_5	$\{(a_8,a_4),(a_5,a_4),(a_3,a_4)\}$	0.5	0.5	0

Table 1: Values of $\texttt{ImpDV}^{\text{Hbs}}_{\text{AS}}(X,y)$ with $\text{AS}_1 = \text{AS} \ominus^{C} Att^-(X)$ and $\text{AS}_2 = \text{AS} \ominus^{\mathcal{A}}_{y} X$.

a_3), the impact of $\{a_1\}$ on a_4 is equal to 0. As for the revised impact measure, it yields 0.015 because the revised measure solves the problem raised in the scenario where "the direct attackers of X impact y via other paths". Removing only the attack from a_2 to a_1 allows the attacks (a_1,a_2), (a_2,a_5) and (a_2,a_3) to be maintained. Hence the acceptability degree of a_4 when removing (a_2,a_1) is different from the acceptability degree of a_4 when removing a_1, resulting in a non-null impact value of $\{a_1\}$ on a_4.

Note that for CS, to limit the problem related to its violation of the Independence property (because of the normalization factor N), we have adopted the same approach as [16], i.e. the same N is used for the two sub-graphs constructed in Definition 10.

3.2 Impact Measure based on Shapley Value

In an argumentative setting, the Shapley contribution measure was used to calculate the contribution of direct attackers on an argument [5]. This approach produced interesting results, but we aim to go beyond this limitation to define an impact measure of any set of arguments (and not just direct attackers) on an argument.

The Shapley measure is a function that associates to each attack a number in $[0,1]$ such that, for each argument, the loss of acceptability is equal to the sum of the values of all the attacks toward it. A gradual semantics satisfies the Attack Removal Monotonicity property if attacks cannot be beneficial for arguments.

Property 3 (Attack Removal Monotonicity [5]). *A semantics σ satisfies Attack Removal Monotonicity iff for any AF $\text{AS} = (\mathcal{A}, C)$, for every $a \in \mathcal{A}$, for every $R \subseteq \{(x,a) \mid x \in Att(a)\}$, it holds that $\sigma_{\text{AS}}(a) \leq \sigma_{\text{AS}\ominus^{C}R}(a)$.*

In [5], they conjectured that h-categoriser satisfies Property 3. We agree with that conjecture and, from our implementations and experiments, we hypothesize that the max-based and card-based semantics also satisfy Property 3. However, due to its nature, we show that the counting semantics does not satisfy Property 3. Indeed, considering the two AFs AS_1 and AS_2 depicted in Figure 1, we see that $\sigma^{\text{CS}}_{\text{AS}_2}(a_3) = 0.26$ is strictly higher than $\sigma^{\text{CS}}_{\text{AS}_1}(a_3) = 0.02$.

Definition 11 (Extended Shapley measure). *Let $\text{AS} = (\mathcal{A}, C)$ be an AF and σ be a gradual semantics. The Shapley measure, w.r.t. σ, is the function $s : C \to [-1,1]$ such that:*

$$s((b,a)) = \sum_{X \subseteq Y} \frac{|X|!(n-|X|-1)!}{n!}(\sigma_{\text{AS}_2}(a) - \sigma_{\text{AS}_1}(a)),$$

where $Y = \{(y,a) \mid y \in Att(a)\} \setminus \{(b,a)\}, n = |Att(a)|, \text{AS}_1 = \text{AS} \ominus^{C} X$, and $\text{AS}_2 = \text{AS} \ominus^{C} (X \cup \{(b,a)\})$.

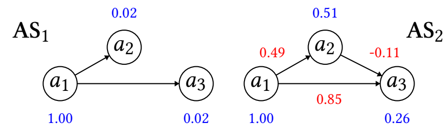


Figure 1: Values in blue represent the acceptability degrees for the counting semantics. Values in red represent the Shapley measure for each of the attacks.

As impact measures must account for a variety of gradual semantics, Definition 11 generalises the Shapley measure of [5] as a function which returns a value in $[-1,1]$ instead of $[0,1]$. Indeed, since the counting semantics does not satisfy Property 3, applying the Shapley measure can result in attacks with negative contributions. Namely, the attacks with negative contributions are attacks that increase the degree of the target argument. In Figure 1 (right), the attack (a_2,a_3) increases the degree of a_3 by 0.11 but (a_1,a_3) decreases the degree of a_3 by 0.85, resulting in a degree of 0.26. This new extended Shapley measure allows to highlight this surprising behavior of gradual semantics which was not possible with the old version. However, note that if a gradual semantics satisfies Attack Removal Monotonicity, then the value returned by the Shapley measure will remain in the interval $[0,1]$.

We now define the Shapley-based impact measure based on this extended Shapley measure. To compute the impact of any set of arguments X on an argument y, the Shapley-based impact measure considers, for each argument x in X, the difference between the contribution of even paths and odd paths from x to y.

Definition 12 (Shapley-based impact measure). *Let $\text{AS} = (\mathcal{A}, C)$ be an AF, $a \in \mathcal{A}, X \subseteq \mathcal{A}, \sigma$ be a gradual semantics, and s be the Shapley measure of Definition 11. The Shapley-based impact measure ImpSI is $\texttt{ImpSI}^{\sigma}_{\text{AS}}(X,a) = \sum_{x \in X} \texttt{ImpSI}^{\sigma}_{\text{AS}}(\{x\},a)$, where:*

$$\texttt{ImpSI}^{\sigma}_{\text{AS}}(\{x\},a) = \left(\sum_{(a_1,\ldots,a_n) \in P_E(x,a)} \prod_{1 \leq i \leq n-1} s((a_{i+1},a_i)) - \sum_{(a_1,\ldots,a_n) \in P_O(x,a)} \prod_{1 \leq i \leq n-1} s((a_{i+1},a_i)) \right)$$

where $P_O(x,a)$ (resp. $P_E(x,a)$) is the set of all odd (resp. even) paths from x to a.

Note that it is possible to obtain a normalized Shapley-based impact measure $\texttt{Imp}'$ such that for all $a \in \mathcal{A}$, $\sum_{a' \in \mathcal{A}} \texttt{Imp}'^{\sigma}_{\text{AS}}(\{a'\},a) = \texttt{Imp}'^{\sigma}_{\text{AS}}(\mathcal{A},a) = 1 - \sigma(a)$, where:
$\texttt{Imp}'(\{a'\},a) = \texttt{ImpSI}^{\sigma}_{\text{AS}}(\{a'\},a)/\texttt{ImpSI}^{\sigma}_{\text{AS}}(\mathcal{A},a) * (1-\sigma(a))$.

The following theorem shows that the Shapley-based impact measure of all arguments in the AF on a given argument will necessarily be negative or neutral.

Theorem 3. *For any gradual semantics σ that satisfies Attack Removal Monotonicity, $\text{AS} = (\mathcal{A}, C)$, and for any $x \in \mathcal{A}$, we have $\texttt{ImpSI}^{\sigma}_{\text{AS}}(\mathcal{A},x) \in [-1,0]$.*

The next property (on gradual semantics) implies that an argument cannot contribute for an absolute Shapley measure that is greater than the degree of the attacker.

Property 4 (Bounded Loss). *We say that a gradual semantics σ satisfies Bounded Loss iff for every $\mathbf{AS} = (\mathcal{A}, C)$, every $a, b \in \mathcal{A}$, $(a, b) \in C$, then $|s((a, b))| \leq \sigma_{\mathbf{AS}}(a)$.*

Conjecture 1. *For any $\sigma \in \{\sigma^{\mathsf{Hbs}}, \sigma^{\mathsf{Car}}, \sigma^{\mathsf{Max}}, \sigma^{\mathsf{CS}}\}$, σ satisfies Bounded Loss.*

Conjecture 2. *Let $\mathbf{AS} = (\mathcal{A}, C)$ be an arbitrary AF. For any gradual semantics σ that satisfies Bounded Loss, it holds that for any $a \in \mathcal{A}$ and $X \subseteq \mathcal{A}$, $\mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(X, a) \in [-1, 1]$.*

The first remark is that when a gradual semantics does not satisfy Property 4, Conjecture 2 is not satisfied as shown by the next example. Let us consider the AF represented in Figure 2 and a gradual semantics σ such that $\sigma_{\mathbf{AS}}(a_2) = 0$ and $\sigma_{\mathbf{AS}}(a_1) = 1/6$, meaning that σ does not satisfy Property 4. We have $\mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a2, a6, a4\}, a1) = \mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a2\}, a1) + \mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a6\}, a1) + \mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a4\}, a1) = -5/6 - 0.5 * 0.5 * 5/6 - 0.5 * 0.5 * 5/6 = -15/12 < -1$.

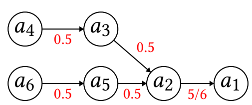


Figure 2: Representation of an AF with 6 arguments. The red number on an edge c represents $s(c)$.

Shapley-based impact measure is not restricted to acyclic graphs. Figure 3 shows the degree of all arguments for the h-categoriser semantics (in blue) and how the Shapley measure associates with each attack in C, its intensity (in red). Using the Shapley-based impact measure (see Definition 12), the impact of $\{a_6\}$ on a_4 is 0.098, of $\{a_5\}$ on a_4 is -0.196, and of $\{a_6, a_5\}$ on a_4 is -0.098. The impact of $\{a_{10}\}$ on a_4 is $\left(\sum_{i=1}^{\infty} (-0.382)^{2(i-1)+1}\right) 0.382 \times 0.235 \simeq -0.0402$.

3.3 Observations

In Table 2, we provide the impact returned by ImpDV and ImpSI of several sets of arguments on a_4 in the AF represented in Figure 3. Since those two approaches follow different intuitions, we can make several observations.

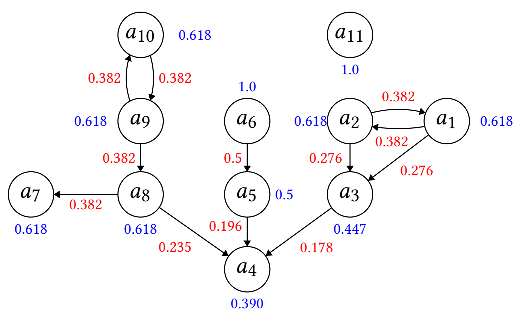


Figure 3: Intensity of attacks with Shapley measure for the h-categoriser semantics. Values in blue are the degrees of the arguments whereas the values in red represent the intensity of the attacks.

	$\mathtt{ImpDV}^{\sigma}_{\mathbf{AS}}(X, a_4)$				$\mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(X, a_4)$			
X	Hbs	Car	Max	CS	Hbs	Car	Max	CS
$\{a_1\}$	0.015	0.001	0.0	0.072	0.036	0.056	0.019	0.026
$\{a_1, a_2\}$	0.069	0.012	0.118	0.161	0.071	0.112	0.037	0.051
$\{a_9\}$	0.069	0.011	0.118	0.107	0.105	0.236	0.061	0.076
$\{a_8\}$	-0.174	-0.082	-0.118	-0.327	-0.235	-0.264	-0.135	-0.291
$\{a_{10}\}$	-0.026	-0.002	-0.018	-0.034	-0.041	-0.138	-0.024	-0.018
$\{a_8, a_{10}\}$	-0.174	-0.082	-0.118	-0.327	-0.276	-0.402	-0.159	-0.309
$\{a_8, a_9, a_{10}\}$	-0.124	-0.072	0.0	-0.246	-0.17	-0.167	-0.098	-0.233
$\{a_5\}$	-0.158	-0.079	-0.118	-0.327	-0.196	-0.255	-0.111	-0.212
$\{a_6\}$	0.064	0.011	0.118	0.107	0.098	0.17	0.056	0.069
$\{a_5, a_6\}$	-0.094	-0.068	0.0	-0.220	-0.098	-0.085	-0.056	-0.143

Table 2: Impacts of several sets of arguments on a_4 using different gradual semantics.

For example, we can see that, for $\sigma \in \{\sigma^{\mathsf{Hbs}}, \sigma^{\mathsf{Car}}, \sigma^{\mathsf{Max}}, \sigma^{\mathsf{CS}}\}$, $\mathtt{ImpDV}^{\sigma}_{\mathbf{AS}}(\{a_8\}, a_4) = \mathtt{ImpDV}^{\sigma}_{\mathbf{AS}}(\{a_8, a_{10}\}, a_4)$ whereas we have that $\mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a_8\}, a_4) \neq \mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a_8, a_{10}\}, a_4)$. The idea is that the Shapley-based impact measure of a set on a target argument can be "decomposed" as the sum of the impact of each argument of that set on the argument. Since $\mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a_{10}\}, a_4) \neq 0$, it holds that $\mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a_8\}, a_4) \neq \mathtt{ImpSI}^{\sigma}_{\mathbf{AS}}(\{a_8, a_{10}\}, a_4)$. In the case of the revised version of ImpDV, the impact of a set X on y is the difference in acceptability degree of y between when the direct external attacks on X are removed and when X is removed. Here, we have that $\sigma_{\mathbf{AS} \ominus^{C} Att^{-}(\{a_8, a_{10}\})}(a_4) = \sigma_{\mathbf{AS} \ominus^{C} Att^{-}(\{a_8\})}(a_4)$ and $\sigma_{\mathbf{AS} \ominus^{\mathcal{A}}_{a_4} \{a_8\}}(a_4) = \sigma_{\mathbf{AS} \ominus^{\mathcal{A}}_{a_4} \{a_8, a_{10}\}}(a_4)$, thus the equality.

For the max-based semantics, the impact of some set of arguments (e.g. $\{a_1\}$ or $\{a_5, a_6\}$) on a_4 is neutral when ImpDV is used. This is not the case when ImpSI as this measure is based on the extended Shapley measure which never assigns a value of 0 to any attacks (for the semantics considered). Apart from these two cases, note that the polarity of the impact (i.e., positive or negative) is often the same for our two approaches, given a set of arguments X and a gradual semantics σ that satisfies Attack Removal Monotonicity.

Our aim is now to provide a principle-based study to compare these impact measures to explain the common features and the differences observed in this subsection.

4 DESIRABLE PRINCIPLES FOR IMPACT MEASURES

Impact measures are usually defined in a general way and can be paired with any gradual semantics (see e.g. [16]). Our aim now is to find out how to evaluate these measures axiomatically. In the case of gradual semantics, a property is satisfied if the result is correct for any AF that satisfies the conditions of the property. If we generalize this to the principles for impact measures, it is obvious that this condition must also be satisfied, but in addition, it would have to hold for any gradual semantics (see Def. 7). However, the challenge arises because there are no constraints on gradual semantics (see Def. 2). Thus, it would be possible to build a particular semantics that violates a principle. That is why, in the rest of this section, we define the desirable principles of a pairing of an impact measure Imp with a gradual semantics σ, denoted $\mathtt{Imp}^{\sigma}$. Namely, $\mathtt{Imp}^{\sigma}$ takes as input any AF AS and returns $\mathtt{Imp}^{\sigma}_{\mathbf{AS}}$. Some of the principles that we introduce and use are inspired by the property analysis of gradual semantics [4], i.e., the Anonymity, Independence, and Directionality properties.

4.1 List of Principles

Unless stated explicitly, all the principles are defined for an impact measure $\texttt{Imp}$ and a gradual semantics σ. Let us start by introducing the notion of isomorphism between argumentation frameworks.

Definition 13. *Let* $\mathbf{AS} = (\mathcal{A}, C)$ *and* $\mathbf{AS}' = (\mathcal{A}', C')$ *be two AFs. An isomorphism from* $\mathbf{AS}$ *to* $\mathbf{AS}'$ *is a bijective function* f *from* $\mathcal{A}$ *to* $\mathcal{A}'$ *such that for all* $a, b \in \mathcal{A}$, $(a, b) \in C$ *iff* $(f(a), f(b)) \in C'$.[2] *If* $\mathbf{AS} = \mathbf{AS}'$, f *is called an automorphism.*

Impact Anonymity states that the impact of a set of arguments on an argument should not depend on the names of the arguments.

Principle 1 (Impact Anonymity). $\texttt{Imp}^\sigma$ *satisfies Impact Anonymity iff for any two AFs* $\mathbf{AS} = (\mathcal{A}, C)$, $\mathbf{AS}' = (\mathcal{A}', C')$, *and any isomorphism* f *from* $\mathbf{AS}$ *to* $\mathbf{AS}'$, *the following holds:* $\forall X \subseteq \mathcal{A}, a \in \mathcal{A}, \texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) = \texttt{Imp}^\sigma_{\mathbf{AS}'}(f(X), f(a))$.

Impact Independence states that the impact of a set of arguments X on an argument a should not depend on the arguments which are not connected to X nor to a by a path.

Principle 2 (Impact Independence). $\texttt{Imp}^\sigma$ *satisfies Impact Independence iff for any two AFs* $\mathbf{AS} = (\mathcal{A}, C)$, $\mathbf{AS}' = (\mathcal{A}', C')$, *where* $\mathcal{A} \cap \mathcal{A}' = \emptyset$, *the following holds:* $\forall X \subseteq \mathcal{A}, a \in \mathcal{A}, \texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) = \texttt{Imp}^\sigma_{\mathbf{AS} \oplus \mathbf{AS}'}(X, a)$.

Balanced Impact states that the sum of the impact of a set of arguments X and the impact of an argument x' on an argument a should be equal to the impact of the union of X with the set containing only argument x' on a. Note that this principle is inspired from [16] but we generalize it to any sets instead of singleton sets. While this principle makes it easier to explain the impact of a set using the impact of its individual elements, it also prevents the modeling of complex behavior such as *accrual*.

Principle 3 (Balanced Impact). $\texttt{Imp}^\sigma$ *satisfies Balanced Impact iff for any AF* $\mathbf{AS} = (\mathcal{A}, C)$, *the following holds:* $\forall X \subseteq \mathcal{A}, x' \in \mathcal{A} \setminus X, a \in \mathcal{A}, \texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) + \texttt{Imp}^\sigma_{\mathbf{AS}}(\{x'\}, a) = \texttt{Imp}^\sigma_{\mathbf{AS}}(X \cup \{x'\}, a)$.

Void Impact states that an empty set has no impact on arguments.

Principle 4 (Void Impact). $\texttt{Imp}^\sigma$ *satisfies Void Impact iff for any AF* $\mathbf{AS} = (\mathcal{A}, C)$, *any* $a \in \mathcal{A}$, $\texttt{Imp}^\sigma_{\mathbf{AS}}(\emptyset, a) = 0$.

Impact Directionality states that the impact of a set of arguments on an argument a remains unchanged when adding an attack in which the target argument is not connected to a via a path.

Principle 5 (Impact Directionality). $\texttt{Imp}^\sigma$ *satisfies Impact Directionality iff for any two AFs* $\mathbf{AS} = (\mathcal{A}, C)$ *and* $\mathbf{AS}' = (\mathcal{A}, C \cup \{(b, x)\})$, *for any* $a \in \mathcal{A}$, *if there is no path from* x *to* a, *then for all* $X \subseteq \mathcal{A}$, $\texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) = \texttt{Imp}^\sigma_{\mathbf{AS}'}(X, a)$.

Impact Minimization captures the fact that the impact of a set of arguments X on an argument a can be reduced to a minimal subset of X from which arguments with no path to a have been removed.

[2] For a function f and the set X, we use the standard notation $f(X)$ to mean $\{f(x) \mid x \in X\}$.

Principle 6 (Impact Minimization). $\texttt{Imp}^\sigma$ *satisfies Impact Minimization iff for any AF* $\mathbf{AS} = (\mathcal{A}, C)$, *any* $X \subseteq \mathcal{A}, x' \in X$ *such that there is no path from* x' *to* a, *and* $a \in \mathcal{A}$, *it holds that* $\texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) = \texttt{Imp}^\sigma_{\mathbf{AS}}(X \setminus \{x'\}, a)$.

Zero Impact states that the impact of an argument x on an argument a is zero if x is not connected to a by a path.

Principle 7 (Zero Impact). $\texttt{Imp}^\sigma$ *satisfies Zero Impact iff for any AF* $\mathbf{AS} = (\mathcal{A}, C)$ *and* $a, x \in \mathcal{A}$, *if there is no path from* x *to* a *then* $\texttt{Imp}^\sigma_{\mathbf{AS}}(\{x\}, a) = 0$.

Impact Symmetry says that if there is an automorphism between the attack structures of two arguments a and b, then the impact of a set of arguments X on a is the same as the impact of the set containing the image of each argument of X on b.

Principle 8 (Impact Symmetry). $\texttt{Imp}^\sigma$ *satisfies Impact Symmetry iff for any AF* $\mathbf{AS} = (\mathcal{A}, C)$, *any* $a, b \in \mathcal{A}$, *the following holds: if* f *is an automorphism from* $\mathbf{AS}|_{\texttt{Str}(a) \cup \texttt{Str}(b)}$ *to* $\mathbf{AS}|_{\texttt{Str}(a) \cup \texttt{Str}(b)}$ *such that* $f(a) = b$ *and* $f(b) = a$, *then for all* $X \subseteq \mathcal{A}, \texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) = \texttt{Imp}^\sigma_{\mathbf{AS}}(f(X \cap (\texttt{Str}(a) \cup \texttt{Str}(b))), b)$.

Example 3. *We illustrate Impact Symmetry with the AF* $\mathbf{AS} = (\mathcal{A}, C)$ *with* $\mathcal{A} = \{a, b, x, y, z\}$ *and* $C = \{(x, a), (y, a), (y, b), (z, b)\}$. *Note that* $\texttt{Str}(a) \cup \texttt{Str}(b) = \mathcal{A}$. *Assume that* $f(a) = b, f(b) = a, f(y) = y, f(x) = z$ *and* $f(z) = x$. *This principle states that* $\texttt{Imp}^\sigma_{\mathbf{AS}}(\{x, y\}, a) = \texttt{Imp}^\sigma_{\mathbf{AS}}(\{y, z\}, b)$.

The next principle states that whenever an argument's final strength differs from 1, there is a set of arguments whose impact explain this difference. This principle is inspired from [26] and ensures that impact measures that always return 0 are not desirable. Without loss of generality, this principle can be easily generalized to gradual semantics which maximal value is not 1.

Principle 9 (Impact Existence). $\texttt{Imp}^\sigma$ *satisfies Impact Existence iff for any AF* $\mathbf{AS} = (\mathcal{A}, C)$, *any* $a \in \mathcal{A}$, *the following holds: if* $\sigma(a) \neq 1$ *then there exists* $X \subseteq A$ *such that* $\texttt{Imp}^\sigma_{\mathbf{AS}}(X, a) \neq 0$.

4.2 Links between Principles

Although most of our principles are independent, some of them are related because they deal with neutral impact where there is no path between the arguments whose impact we want to calculate and the target argument. This is the case with Impact Symmetry and Impact Minimization which follows from some other principles.

Theorem 4. *Let* $\texttt{Imp}$ *be an impact measure and* σ *be a gradual semantics. If* $\texttt{Imp}^\sigma$ *satisfies Impact Anonymity, Impact Directionality, Impact Minimization, and Impact Independence, then* $\texttt{Imp}^\sigma$ *satisfies Impact Symmetry.*

Theorem 5. *Let* $\texttt{Imp}$ *be an impact measure and* σ *be a gradual semantics. If* $\texttt{Imp}^\sigma$ *satisfies Zero Impact and Balanced Impact, then* $\texttt{Imp}^\sigma$ *satisfies Impact Minimization.*

Theorem 6. *Let* $\texttt{Imp}$ *be an impact measure and* σ *be a gradual semantics. If* $\texttt{Imp}^\sigma$ *satisfies Void Impact and Impact Minimization, then* $\texttt{Imp}^\sigma$ *satisfies Zero Impact.*

Defining principles solely at the level of impact measure is not relevant, as the calculation is based on the use of scores returned by gradual semantics. A number of axiomatic studies have been carried out on these semantics to guarantee correct behavior and better understand the results returned. In this way, it is possible to link the satisfaction of some properties to the satisfaction of other principles for $\texttt{ImpDV}^\sigma$ and $\texttt{ImpSI}^\sigma$.

Proposition 7. *$\texttt{ImpSI}^\sigma$ and $\texttt{ImpDV}^\sigma$ satisfy Impact Independence for any gradual semantics σ that satisfies Independence.*

Proposition 8. *$\texttt{ImpSI}^\sigma$ and $\texttt{ImpDV}^\sigma$ satisfy Impact Directionality for any gradual semantics σ that satisfies Directionality.*

Proposition 9. *$\texttt{ImpDV}^\sigma$ satisfies Zero Impact for any gradual semantics σ that satisfies Directionality and Independence.*

Proposition 10. *If $\texttt{ImpDV}^\sigma$ satisfies Impact Directionality, then it satisfies Impact Minimization for any gradual semantics σ that satisfies Independence.*

4.3 Axiomatic Evaluation and Discussion

Table 3 summarizes the results of our the axiomatic evaluation. This axiomatic study only takes into consideration the gradual semantics defined in Section 2. This choice is motivated by the fact that some of these semantics have already been studied both axiomatically (and in association with an existing impact measure), and others have only been studied axiomatically but have unique features (e.g., the use of the max aggregation) that were interesting to study. While there exist other gradual semantics that could have been studied, the idea was to show that, despite the different semantics' behaviors, our impact measures can be applied to any gradual semantics.

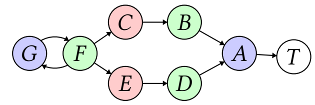
Theorem 11. *The principles of Table 3 hold.*

We can observe that $\texttt{ImpDV}^\sigma$ and $\texttt{ImpSI}^\sigma$ both satisfy Impact Anonymity, Void Impact, Impact Minimization, Zero Impact and Impact Symmetry for $\sigma \in \{\sigma^{\texttt{Hbs}}, \sigma^{\texttt{Car}}, \sigma^{\texttt{Max}}, \sigma^{\texttt{CS}}\}$.

The principles that are not satisfied by $\texttt{ImpDV}^\sigma$ and $\texttt{ImpSI}^\sigma$ include Impact Independence and Impact Directionality only for CS. This can be explained by the behavior of CS that does not satisfy the Independence and the Directionality properties. The main axiomatic difference between the two measures concerns the Balanced Impact principle because ImpSI satisfies this principle whatever the gradual semantics used in our study, whereas this is never the case for ImpDV. The AF depicted in Figure 3 shows, for example, that $\texttt{ImpDV}^\sigma_{\texttt{AS}}(\{a_8, a_{10}\}, a_4) \neq \texttt{ImpDV}^\sigma_{\texttt{AS}}(\{a_8\}, a_4) + \texttt{ImpDV}^\sigma_{\texttt{AS}}(\{a_{10}\}, a_4)$ for $\sigma \in \{\sigma^{\texttt{Hbs}}, \sigma^{\texttt{Car}}, \sigma^{\texttt{Max}}, \sigma^{\texttt{CS}}\}$. Note also that, although [16, Proposition 2] states that Balanced Impact is satisfied by the original definition of $\texttt{ImpDV}^{\sigma^{\texttt{CS}}}$ (cf. Definition 9), this AF can also be used as a counterexample to prove that it is not satisfied.

Moreover, while most of the proofs have been done on general graphs, we have proved the satisfaction of Impact Existence for $\texttt{ImpSI}^{\sigma^{\texttt{CS}}}$ only on the class of graphs where there are at least two arguments with the maximum in-degree, however we conjecture that Impact Existence is also satisfied in the general case.

This principle compliance study can be helpful for choosing which (impact, semantics) pair to use for a specified application. It is interesting to note here that this choice depends on two factors: The satisfiability of the Balanced Impact principle and the satisfiability of the Impact Independence and the Impact Directionality principles.



	$\texttt{ImpDV}^\sigma_{\texttt{AS}}(X,T)$		$\texttt{ImpSI}^\sigma_{\texttt{AS}}(X,T)$	
X	Hbs	CS	Hbs	CS
$Ag_1 - \{A, G\}$	-0.5	-0.49	-0.32	-0.172
$Ag_2 - \{C, E\}$	-0.08	-0.235	-0.065	-0.036
$Ag_3 - \{B, D, F\}$	0.25	0.480	0.2	0.124
env $- \{A, G\}$	-0.5	-0.49	-0.32	-0.172
soc $- \{B, C\}$	0.061	0.122	0.053	0.037
eco $- \{F\}$	0.033	0.115	0.029	0.013
infra $- \{D, E\}$	0.061	0.122	0.053	0.037

Figure 4: Argumentation graph of the case study and a table containing the impact of each set of arguments studied on T.

If an application demands that all principles should be satisfied, then we would use ImpSI with one of the three following semantics (Hbs, Max, Car). If an application demands the use of ImpDV then we know that, depending on the semantics we choose, we will have Impact Independence and Impact Directionality satisfied (or not). Finally, if an application demands the use of the counting semantics, then we know that for both impact measures, the Impact Independence and the Impact Directionality principles are not satisfied while Balanced Impact is satisfied only by ImpSI.

5 ILLUSTRATION SCENARIO

To illustrate the impact measures in practice, we present an example inspired by [1] on which we evaluate, among other things, the impact of each agent in a debate about pollution becoming a major health problem in big cities.

Consider the following topic T: "Polluting vehicles, and specifically diesel cars should be maintained in the city centres". A city council composed of three agents might entertain the following arguments (see Figure 4):

- A [Ag_1] Diesel cars should be banned from the inner city center in order to decrease pollution.
- B [Ag_3] Artisans, who deserve special protection by the city council, cannot change their vehicles, as that would be too expensive for them.
- C [Ag_2] The city can offer financial assistance to artisans.
- D [Ag_3] There are only very few alternatives to using diesel cars. Specifically, the autonomy of electric cars is poor, as there are not enough charging stations around.
- E [Ag_2] The city can set up more charging stations.
- F [Ag_3] In times of financial crisis, the city should not commit to spending additional money.
- G [Ag_1] Health and climate change issues are important, so the city has to spend what is needed to tackle pollution.

Knowing the impact of an agent clearly requires calculating the impact of a set of arguments on the topic of the debate T. Thus, we apply our two impact measures coupled with the semantics Hbs and CS to calculate the impact of all these subsets (see Figure 4). Let us note that for both approaches and both semantics, Ag_1 is the agent

Principle	$\texttt{ImpDV}^{\texttt{Hbs}}$	$\texttt{ImpSI}^{\texttt{Hbs}}$	$\texttt{ImpDV}^{\texttt{Max}}$	$\texttt{ImpSI}^{\texttt{Max}}$	$\texttt{ImpDV}^{\texttt{Car}}$	$\texttt{ImpSI}^{\texttt{Car}}$	$\texttt{ImpDV}^{\texttt{CS}}$	$\texttt{ImpSI}^{\texttt{CS}}$
Impact Anonymity	✓	✓	✓	✓	✓	✓	✓	✓
Impact Independence	✓	✓	✓	✓	✓	✓	✗	✗
Balanced Impact	✗	✓	✗	✓	✗	✓	✗	✓
Void Impact	✓	✓	✓	✓	✓	✓	✓	✓
Impact Directionality	✓	✓	✓	✓	✓	✓	✗	✗
Impact Minimization	✓	✓	✓	✓	✓	✓	✓	✓
Zero Impact	✓	✓	✓	✓	✓	✓	✓	✓
Impact Symmetry	✓	✓	✓	✓	✓	✓	✓	✓
Impact Existence	✓	✓	✓	✓	✓	✓	✓	$✓^{\geq 2}$

Table 3: Properties satisfied by the studied impact measures and gradual semantics. The symbol ✓(resp. ✗) means that the property is satisfied (resp. violated) by the impact measure and the gradual semantics considered. The symbol $✓^{\geq 2}$ means that the property is satisfied on the class of graphs where there are at least two arguments with the maximum in-degree.

with the greatest (negative) impact, clearly showing he is against maintaining these cars. Conversely, Ag_3 has a positive impact on the issue of the debate, but is less important than Ag_1. Finally, Ag_2 has little impact (except for $\texttt{ImpDV}^{\texttt{CS}}$). In addition to the sets of arguments proposed by the agents, we can associate the arguments presented in this example with four types of values. Arguments A and G concern environmental responsibility (value env), B and C are about social fairness (value soc), F promotes economic viability (value econ), and D and E pertain to infrastructure efficiency (value infra). In this case, the set of arguments env has the greatest (negative) impact, while all the other categories have a positive but much smaller impact.

6 RELATED WORK

Yin et al. [41] introduced an impact measure to explain the Discontinuity Free Quantitative Argumentation Debate (DF-QuAD) gradual semantics [35] in quantitative bipolar argumentation frameworks (QBAFs). Their impact measure quantifies the contribution of an argument towards topic arguments in QBAFs. Although their work is also inspired by feature attribution explanation methods in machine learning, Yin et al. focus on highlighting the sensitivity of a topic argument's final acceptability degree w.r.t. the other arguments' initial weights. Their impact measure is defined only for individual arguments, in acyclic QBAFs and only for the DF-QuAD semantics. However, our two impact measures are both defined for any set of arguments and can be paired with any gradual semantics. Moreover, the properties they study are explanation-focused, used to assess and characterize their impact measure's ability of providing robust and faithful explanations. These properties are mostly inspired by properties for machine learning models' explanations such as sensitivity and fidelity. Here, we propose contribution-focused properties, meaning that we evaluate how each pair (impact, semantics) contributes to the final acceptability degree of an argument, w.r.t. the argumentation graph's structure. However, we also intend to explore the explanation-focused properties. Namely, we want to study how to produce "good" explanations for gradual semantics using the impact measures that we defined in this paper.

Kampik et al. [26] propose *contribution functions* and principles in the context of quantitative bipolar argumentation. Contrary to our work, their contribution functions are only defined for acyclic graphs and only measure the influence of a single source argument on a topic argument. While they also introduce a Shapley-based contribution, its computation necessitates the addition of the source argument to all possible sub-graphs that already contain the topic argument. We argue that this is more computationally expensive than our Shapley-based impact measure which is based on [5].

The notion of impact for gradual semantics has also been studied by Himeur et al. [24]. They measured the impact of agents on arguments in a debate. Although the impact measure defined returns the individual impact of an argument on another argument, they defined different aggregation functions that can be used to merge the impact of all the arguments belonging to the same agent, on a particular argument. The impact measure of Himeur et al. [24] shares similarities with to the one defined in [16]. Moreover, their impact measure is studied only for Euler-based semantics [3] and DF-QuAD w.r.t. a set of principles and aggregation functions.

7 CONCLUSION AND FUTURE WORK

We studied the notion of impact of a set of arguments on an argument under gradual semantics. We proposed two impact measures: ImpDV, a revision of the measure from [16], and ImpSI, a novel measure based on the Shapley Contribution Measure, which is derived from the measure introduced in [5]. We provided a principle-based analysis of these two impact measures under four semantics: h-categoriser (Hbs), card-based (Car), max-based (Max) and counting semantics (CS). For three of the gradual semantics (Hbs, Max, Car), we show that $\texttt{ImpSI}^{\sigma}$ satisfies all the principles, while $\texttt{ImpDV}^{\sigma}$ satisfies them all except Balanced Impact. Concerning CS, our two impact measures do not satisfy Impact Independence and Impact Directionality because the two associated properties for gradual semantics are not satisfied by CS. For future work, we plan to study how to generate explanations for gradual semantics based on these two impact measures. Providing explanations in abstract AFs has been explored for extension-based semantics [9, 13, 14, 18–21, 27, 38]. However, it has not been explored for gradual semantics. Our results allow us to study how we can use impact measures to provide explanations for gradual semantics outcomes.

ACKNOWLEDGMENTS

This work benefited from the support of the project AGGREEY ANR-22-CE23-0005 of the French National Research Agency (ANR).

REFERENCES

[1] Stéphane Airiau, Elise Bonzon, Ulle Endriss, Nicolas Maudet, and Julien Rossit. 2016. Rationalisation of Profiles of Abstract Argumentation Frameworks. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, Singapore, May 9-13, 2016*, Catholijn M. Jonker, Stacy Marsella, John Thangarajah, and Karl Tuyls (Eds.). ACM, 350–357. http://dl.acm.org/citation.cfm?id=2936976

[2] Leila Amgoud and Jonathan Ben-Naim. 2013. Ranking-Based Semantics for Argumentation Frameworks. In *Scalable Uncertainty Management - 7th International Conference, SUM 2013, Washington, DC, USA, September 16-18, 2013. Proceedings (Lecture Notes in Computer Science, Vol. 8078)*, Weiru Liu, V. S. Subrahmanian, and Jef Wijsen (Eds.). Springer, 134–147. https://doi.org/10.1007/978-3-642-40381-1_11

[3] Leila Amgoud and Jonathan Ben-Naim. 2018. Evaluation of arguments in weighted bipolar graphs. *International Journal of Approximate Reasoning* 99 (2018), 39–55.

[4] Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, and Srdjan Vesic. 2017. Acceptability Semantics for Weighted Argumentation Frameworks.. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. 56–62. https://doi.org/10.24963/ijcai.2017/9

[5] Leila Amgoud, Jonathan Ben-Naim, and Srdjan Vesic. 2017. Measuring the Intensity of Attacks in Argumentation Graphs with Shapley Value.. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. 63–69. https://doi.org/10.24963/ijcai.2017/10

[6] Leila Amgoud, Yannis Dimopoulos, and Pavlos Moraitis. 2007. A unified and general framework for argumentation-based negotiation. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. 1–8.

[7] Leila Amgoud, Dragan Doder, and Srdjan Vesic. 2022. Evaluation of argument strength in attack graphs: Foundations and semantics. *Artificial Intelligence* 302 (Jan. 2022), 103607. https://doi.org/10.1016/j.artint.2021.103607

[8] Leila Amgoud and Henri Prade. 2009. Using arguments for making and explaining decisions. *Artificial Intelligence* 173, 3-4 (2009), 413–436.

[9] Ringo Baumann and Markus Ulbricht. 2021. Choices and their Consequences-Explaining Acceptable Sets in Abstract Argumentation Frameworks.. In *KR*. 110–119.

[10] Philippe Besnard and Anthony Hunter. 2001. A logic-based theory of deductive arguments. *Artif. Intell.* 128, 1-2 (2001), 203–235. https://doi.org/10.1016/S0004-3702(01)00071-6

[11] Vivien Beuselinck, Jérôme Delobelle, and Srdjan Vesic. 2023. A Principle-based Account of Self-attacking Arguments in Gradual Semantics. *J. Log. Comput.* 33, 2 (2023), 230–256. https://doi.org/10.1093/LOGCOM/EXAC093

[12] Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. 2023. An empirical and axiomatic comparison of ranking-based semantics for abstract argumentation. *J. Appl. Non Class. Logics* 33, 3-4 (2023), 328–386. https://doi.org/10.1080/11663081.2023.2246863

[13] AnneMarie Borg and Floris Bex. 2021. A basic framework for explanations in argumentation. *IEEE Intelligent Systems* 36, 2 (2021), 25–35.

[14] AnneMarie Borg and Floris Bex. 2022. Contrastive Explanations for Argumentation-Based Conclusions. In *21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022*, Piotr Faliszewski, Viviana Mascardi, Catherine Pelachaud, and Matthew E. Taylor (Eds.). International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 1551–1553. https://doi.org/10.5555/3535850.3536031

[15] Jérôme Delobelle. 2017. *Ranking-based Semantics for Abstract Argumentation. (Sémantique à base de Classement pour l'Argumentation Abstraite)*. Ph.D. Dissertation. Artois University, Arras, France. https://tel.archives-ouvertes.fr/tel-01937279

[16] Jérôme Delobelle and Serena Villata. 2019. Interpretability of Gradual Semantics in Abstract Argumentation. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 15th European Conference, ECSQARU 2019, Belgrade, Serbia, September 18-20, 2019, Proceedings (Lecture Notes in Computer Science, Vol. 11726)*, Gabriele Kern-Isberner and Zoran Ognjanovic (Eds.). Springer, 27–38. https://doi.org/10.1007/978-3-030-29765-7_3

[17] Phan Minh Dung. 1995. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artif. Intell.* 77, 2 (1995), 321–358. https://doi.org/10.1016/0004-3702(94)00041-X

[18] Phan Minh Dung, Robert A Kowalski, and Francesca Toni. 2009. Assumption-based argumentation. *Argumentation in artificial intelligence* (2009), 199–218.

[19] Phan Minh Dung, Paolo Mancarella, and Francesca Toni. 2007. Computing ideal sceptical argumentation. *Artificial Intelligence* 171, 10-15 (2007), 642–674.

[20] Xiuyi Fan and Francesca Toni. 2015. On computing explanations in argumentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.

[21] Xiuyi Fan and Francesca Toni. 2015. On explanations for non-acceptable arguments. In *Theory and Applications of Formal Argumentation: Third International Workshop, TAFA 2015, Buenos Aires, Argentina, July 25-26, 2015, Revised Selected Papers 3*. Springer, 112–127.

[22] Yang Gao, Francesca Toni, Hao Wang, and Fanjiang Xu. 2016. Argumentation-based multi-agent decision making with privacy preserved. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. 1153–1161.

[23] Taehyun Ha, Sangwon Lee, and Sangyeon Kim. 2018. Designing explainability of an artificial intelligence system. In *Proceedings of the Technology, Mind, and Society*. 1–1.

[24] Areski Himeur, Bruno Yun, Pierre Bisquert, and Madalina Croitoru. 2021. Assessing the Impact of Agents in Weighted Bipolar Argumentation Frameworks. In *Artificial Intelligence XXXVIII - 41st SGAI International Conference on Artificial Intelligence, AI 2021, Cambridge, UK, December 14-16, 2021, Proceedings (Lecture Notes in Computer Science, Vol. 13101)*, Max Bramer and Richard Ellis (Eds.). Springer, 75–88. https://doi.org/10.1007/978-3-030-91100-3_6

[25] Takayuki Ito and Toramatsu Shintani. 1997. Persuasion among Agents: An Approach to Implementing a Group Decision System Based on Multi-Agent Negotiation. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, IJCAI 97, Nagoya, Japan, August 23-29, 1997, 2 Volumes*. Morgan Kaufmann, 592–599.

[26] Timotheus Kampik, Nico Potyka, Xiang Yin, Kristijonas Cyras, and Francesca Toni. 2024. Contribution functions for quantitative bipolar argumentation graphs: A principle-based analysis. *Int. J. Approx. Reason.* 173 (2024), 109255. https://doi.org/10.1016/J.IJAR.2024.109255

[27] Beishui Liao and Leendert Van Der Torre. 2020. Explanation semantics for abstract argumentation. In *Computational Models of Argument*. IOS Press, 271–282.

[28] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[29] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[30] Jann Müller and Anthony Hunter. 2012. An argumentation-based approach for decision making. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, Vol. 1. IEEE, 564–571.

[31] Nir Oren, Bruno Yun, Srdjan Vesic, and Murilo S. Baptista. 2022. Inverse Problems for Gradual Semantics. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, 2719–2725. https://doi.org/10.24963/ijcai.2022/377

[32] Alison R Panisson. 2017. Argumentation schemes and enthymemes in multi-agent systems. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 1849–1850.

[33] Nico Potyka. 2018. Continuous dynamical systems for weighted bipolar argumentation. In *Sixteenth International Conference on Principles of Knowledge Representation and Reasoning*.

[34] Fuan Pu, Jian Luo, Yulai Zhang, and Guiming Luo. 2015. Attacker and Defender Counting Approach for Abstract Argumentation. In *Proc. of the 37th Annual Meeting of the Cognitive Science Society, (CogSci'15)*.

[35] Antonio Rago, Francesca Toni, Marco Aurisicchio, and Pietro Baroni. 2016. Discontinuity-Free Decision Support with Quantitative Argumentation Debates.. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016*. 63–73. http://www.aaai.org/ocs/index.php/KR/KR16/paper/view/12874

[36] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[37] Ariel Rosenfeld and Sarit Kraus. 2016. Strategical argumentative agent for human persuasion. In *ECAI 2016*. IOS Press, 320–328.

[38] Zeynep G Saribatur, Johannes P Wallner, and Stefan Woltran. 2020. Explaining non-acceptability in abstract argumentation. In *ECAI 2020*. IOS Press, 881–888.

[39] LS Shapley. 1953. A value for n-person games. *Contributions to the theory of games* 2 (1953), 307–317.

[40] Giulia Vilone and Luca Longo. 2021. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion* 76 (2021), 89–106.

[41] Xiang Yin, Nico Potyka, and Francesca Toni. 2023. Argument Attribution Explanations in Quantitative Bipolar Argumentation Frameworks. In *ECAI 2023 - 26th European Conference on Artificial Intelligence (Frontiers in Artificial Intelligence and Applications, Vol. 372)*, Kobi Gal, Ann Nowé, Grzegorz J. Nalepa, Roy Fairstein, and Roxana Radulescu (Eds.). IOS Press, 2898–2905.